DOCUMENT RESUME

ED 168 458 IR 005 310

AUTHOR Crooks, R. E.
TITLE The Evaluation of an Information Retrieval System.
PUB DATE 74
NOTE 19p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Bibliographies; Cost Effectiveness; Flow Charts; *Information Retrieval; *Literature Reviews; Mathematical Models; *Models; *On Line Systems; *Performance Criteria; Relevance (Information Retrieval); Systems Approach; User Satisfaction (Information)

ABSTRACT

Various models that have been devised to evaluate the performance of information retrieval systems are reviewed. Such models fall into three categories: (1) evaluation from the viewpoint of the user, who is primarily concerned with how well the system responds to his query; (2) evaluation from the system operator's standpoint, which focuses on the operation of the system; and (3) a merging of these two points of view. Parameters considered include system costs, recall, precision, form of output, and response time. Although most of the models described are mathematically derived, no attempt is made to elaborate upon the equations. A bibliography is attached. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED168458

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE EVALUATION OF AN INFORMATION RETRIEVAL SYSTEM

R. E. Crooks

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.E. Crooks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

1 R005310

THE EVALUATION OF AN INFORMATION RETRIEVAL SYSTEM

The purpose of this paper is to describe some of the various ways that experts in the field of Information Science have attempted to create and define models which will evaluate the performance of an information retrieval system. Lancaster (10) defines information retrieval systems as systems that retrieve documents or document references. The approaches to this knotty problem of "how can we evaluate an information retrieval system's performance" fall basically into three categories: (1) the evaluation of the information retrieval system from the user's requirements; (2) the evaluation of the information retrieval system from the system operator's standpoint--which is primarily concerned with the operation of the system; and (3) the evaluation of an information retrieval system resulting from the merging of the user's requirements with those of the system operator's standpoint. All approaches have, however, used the model approach to evaluate the system. A model as used in this paper can be defined as a framework for merging evaluation measures so that their implications can be seen as a unit and one may analyze the utility of the various aspects of the system with respect to whatever needs may be specified. Williams (15) states that the function of a model is to describe the parameters of an information system and to provide measurement and control guides. Most models discussed in this paper are mathematically derived and no attempt has been made to elaborate upon or describe the particular mathematical equations.

First to be discussed is the category of evaluating an information retrieval system from the system operator's needs and costs. Rettenmayer (13) has stated that most attempts to improve the performance of information retrieval systems have concentrated primarily on devising optimal schemes of indexing a file.

Keith (7) has discussed the parameters for a general evaluation model and favors a systems approach. His parameters are the major operational characteristics of a system and their costs related to the user's information storage and retrieval requirements. By the systems approach, all aspects of a storage and retrieval system can be analyzed. Then, by simply comparing systems, one can see which gives the least overall cost in comparison with performance. His formula for accomplishing this goal is:

total expected cost (C_t) = initialization cost $E(C_i)$ +
maintenance cost $E(C_m)$ +
input cost $E(C_{in})$ +
computer operation cost $E(C_{op})$ +
output cost $E(C_{out})$

The $E(C_i)$ or initialization cost can be broken down into the costs for the implementation of the information storage and retrieval system, staffing expenses, support equipment, and the expenses of converting and processing the records to machine readable form and may be expressed in the mathematical form:

$$E(C_i) = E(C_{imp}) + E(C_{st}) + E(C_{eq}) + E(C_{con})$$

The $E(C_m)$ or maintenance cost is determined from the cost to keep the information file up-to-date and accurate, the estimated cost of editing once the frequency of use of the information file has been determined, and

the amount of information that is to be put into the file has been estimated. This may be mathematically expressed as:

$$E(C_m) = E(C_{upd}) + E(C_{edt})$$

$E(C_{in})$ or input cost is the expense of preparing a query and processing that query into the computer and may be expressed mathematically as:

$$E(C_{in}) = E(C_{qpr}) + E(C_{pro})$$

$E(C_{op})$ or computer operation cost may be determined by the amount of time that a query takes to be run and the cost to the user because of a delay in receiving query results. This may be expressed mathematically as:

$$E(C_{op}) = E(C_{opr}) + E(C_{dly})$$

$E(C_{out})$ or output cost may be determined by such factors as complexity of output headings, number of pages of output, the output format, and the time to list query results and return the results to the user. Expressed mathematically, this is:

$$E(C_{out}) = E(C_{for}) + E(C_{list})$$

In any event, though, one must consider the operational characteristics of the system relative to the information needs and most assuredly, one must compromise on some characteristic of the system, i.e. make tradeoffs, if a realistic and useful system is to be the end product.

Cooper (3) evaluates the information retrieval system in terms of cost also, but his model is not as mathematically orientated to the layman as Keith's. In the past models were designed to evaluate the performance of an information retrieval system by using a technique of comparing query representation with surrogates and Cooper maintains that this does not ensure the "best" operation. Cooper's cost model divides the activity of the retrieval system into two parts--the allocation of effort for a given

search between the user and the system.

The user's costs include the cost per unit of the user's time and the amount of his time expended.

The total time of an interaction is divided into stages: (1) presearch; (2) search; and (3) post search.

System costs during the pre search stage include the cost of computer channels, central processing unit costs, and the cost per amount of core storage per unit of time. In the search stage, system costs are a function of the search comparison method and the type of surrogation used in document representation. Post search system costs are also a function of the search comparison method and also the number of terms in the query, the number of logical operators in the query, the number of surrogates in the file, and the number of words in each surrogate. In addition one must calculate the costs for the surrogates, the creation of the system, storage costs, and processing costs.

Figure 1 (see page 5) represents the stages in retrieval as seen from both the user and the system and the activities of each in every stage.

Cigler (2) proposes a simulation of an information process to give feedback on the system's performance. This method of simulation results in a reduction of time and a much lower cost (to evaluate the performance of a retrieval system). To simulate, one must first get the characteristics of the retrieval system and approximate its performance. One must also determine the identification of ways for the processing of retrieval searches in the system; the time consumption of the system; the estimated number of requests; and a time reserve for down or maintenance time.

	User Activities	System Activities
Pre search	1. Determine information need	1. Syntax check of query
	2. Enter query	2. Thesaurus lookup
	3. Query negotiation	3. Query term frequency analysis
		4. Map query into formal language
Search	1. Wait	1. Select comparison method
		2. Search
Post search	1. Read output	1. Display output
	2. Mark relevant material	2. Calculate performance measure
	3. Use relevant material	3. Revise strategies &/or query with feedback

Fig. I

Three Stages and Activities in Retrieval
(3:307f)

Figure II (see page 6) represents Cigler's proposed simulation model for simulating the information process.

Models may also be designed from the standpoint of the user's requirements. Papier (12) has stated that the ability of a system to meet the user's requirements is the prime measure of the effectiveness of a system. Rzasa and Baker (14) in an attempt to measure the effectiveness of university

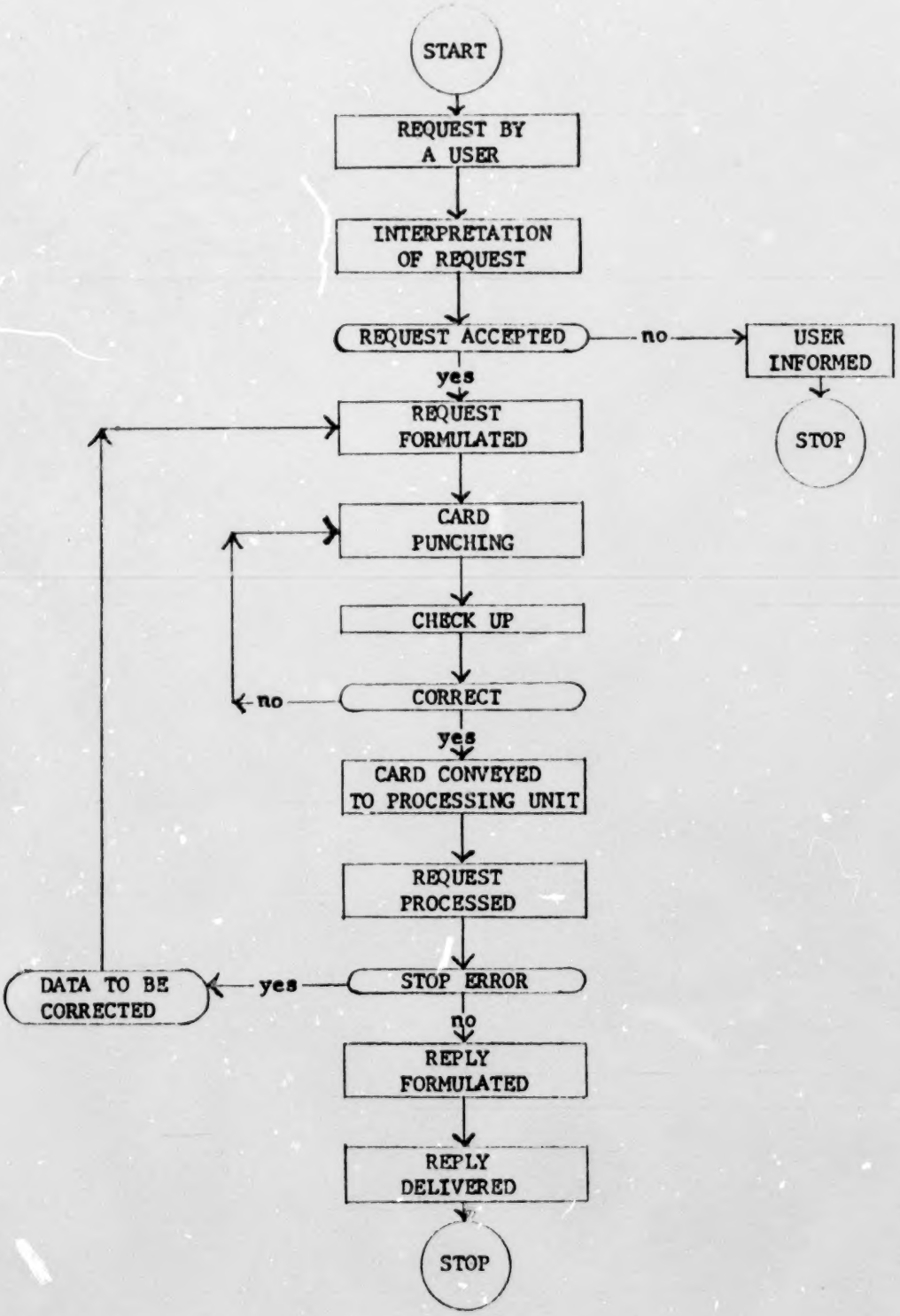
START
REQUEST BY A USER
INTERPRETATION OF REQUEST
REQUEST ACCEPTED
no
USER INFORMED
STOP
yes
REQUEST FORMULATED
CARD PUNCHING
CHECK UP
no
CORRECT
yes
CARD CONVEYED TO PROCESSING UNIT
REQUEST PROCESSED
DATA TO BE CORRECTED
yes
STOP ERROR
no
REPLY FORMULATED
REPLY DELIVERED
STOP

Fig. II (2)

libraries developed two measures of effectiveness. The first is the Primary Measure of Effectiveness and consists of (1) the total number of users and the total number of material items used, (2) the total number of non-material items sought and the total number supplied, and (3) the total number of users who are studying with their own materials or who are at the library for social reasons. By determining the numbers which apply to each criteria, one is able to obtain a rough approximation of the library's effectiveness. They then divided the total population of the university into subpopulations and developed another measure--the Secondary Measure of Effectiveness. This consists of (1) the number of users from a subpopulation (faculty, graduate students, undergraduate students, staff and others) and (2) the return rate of users. These measures, when taken in totality or entirety, give an efficient assessment of the effectiveness of a library.

This viewpoint of evaluating a retrieval system from the user's requirements is typified by Hjerppe and Lindqvist (6). They believe that an information retrieval system can be evaluated in terms of two performance parameters, recall and precision, and that the system should be evaluated from the user's view which includes such factors as his information needs and uses and communication theory. In summary, the evaluation should be result oriented since it is believed that the objective of any information system is to deliver the right amount of information in the right form at the right time. In order to make a decision concerning a specific system, several parameters are delineated and they include: (1) time--the amount of time available for a decision; (2) the quality requirements which are a function of goals, policies, etc.; (3) resources--budget available; (4) previous knowledge--prior knowledge before use of system; and (5) frequency

type--one time or an ongoing requirement. Parameters for the information system are also delineated and include five classes:

1. Coverage related parameters include:
    a. Exhaustivity--the extent to which the knowledge in the field is covered;
    b. Depth--the extent to which a specific question is covered;
    c. Width--the interdisciplinary coverage; and
    d. Retrospectivity--the time coverage of the files.
2. Input related parameters include:
    a. The effort required for the user to learn how to use the system;
    b. The effort required for the formulation of the query; and
    c. The effort required for any modification, if necessary, of the query.
3. Performance related parameters include:
    a. The response time from the first input to the first output;
    b. The channel width--the amount of information received from the system compared to the total amount of information presented to the user;
    c. Ranking--the output sorted in relevance order;
    d. Reliability--the uptime;
    e. Adaptability--the ability to modify a query; and
    f. Sensitivity--instant adaptability.
4. Output related parameters include:
    a. Actuality of output--the time lag of the output;
    b. Presentation--the readability of the output;
    c. Bibliographic level--the content of the output; and
    d. Precision--related to readability and includes the effort of the user to select needed information from the output.

5. Parameters related to the information services as a whole include:

   a. Organization--the service's affiliation;

   b. Privacy--the protection of the input/output; and

   c. Price of the service.

Hjerppe and Lindqvist then propose that a matrix be drawn (Fig. III--see page 10) in which one may see the interaction between the decision parameters and those of the information retrieval system.

Cooper (4) has proposed a simulation model to evaluate a retrieval system from a viewpoint which merges both a system operator's and a user's point of view. It is proposed that one gets a collection of documents, converts them to machine readable form, and poses the queries which are then converted to machine readable form. It is then possible to evaluate specific rules (matching, searching, associative searching) to see which one produces the better results. The performance of the system is measured in terms of recall and precision ratios, but the user is the one who makes the required judgments as to the effectiveness of the system. Cooper believes that this simulation model would measure the effect of changes in the characteristics of the query and document files on the quantity of material received by the user.

King and Bryant (8,9) feel that due to the complex nature of the retrieval system that an evaluation of any retrieval system should be centered on the specific system processes, such as search strategies, indexing, and the interface of the user and the system. Figure IV (see page 11) shows the relationships between the system and the user and can be utilized to evaluate the accuracy of the system or to diagnose any failure in it.

	Time Available	Resources	Previous Knowledge	Frequency Type	Quality Requirements
I. Coverage related					
1. exhaustivity			X		X
2. depth					
3. width					
4. retrospectivity					X
II. Input related					
1. effort required to learn how to use system	X				
2. effort required for formulation of query	X				
3. effort required for modification of query				X	
III. Performance related					
1. response time	X				
2. channel width					X
3. ranking	X				
4. reliability					X
5. adaptability				X	
6. sensitivity				X	X
IV. Output related					
1. actuality of output					X
2. presentation			X		
3. bibliographic level	X	X			
4. precision	X	X			
V. Parameters of Information Services					
1. organization					
2. privacy					
3. price		X			

Fig. III MATRIX of Decision Parameters and Information Retrieval System Parameters (6)

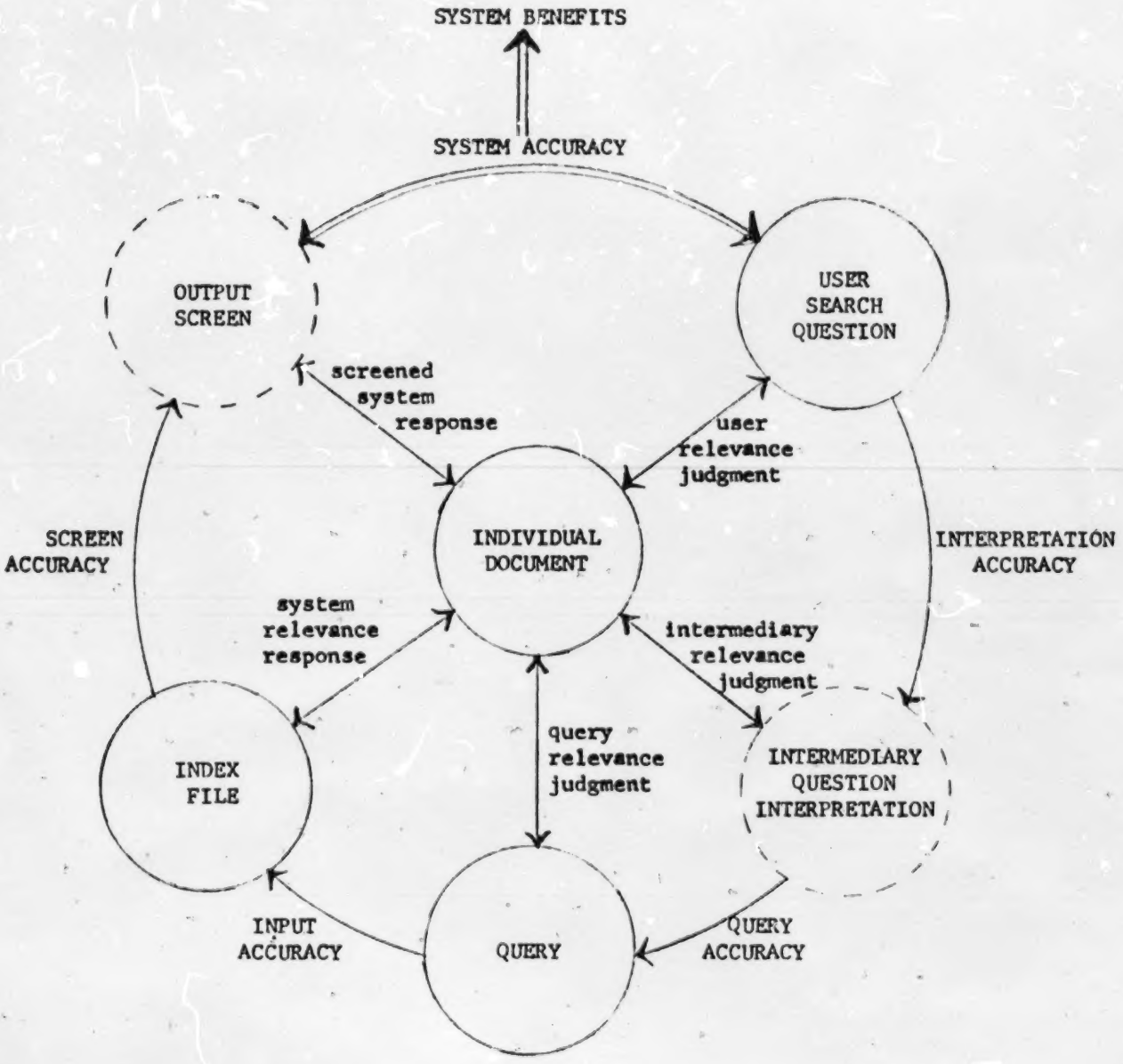

Fig. IV

System Accuracy and

Relationships That Contribute To It

(8:263f)

The user relevance judgment is the relationship between the search question and the document.

The intermediary relevance judgment is the judgment of degree of relationship between the query and the document.

The query relevance judgment is the assessment of degree of relationship between the query and the individual document.

The system relevance response is the system's assessment of the relationship between the query and the individual document.

The screened system relevance response is an intermediary's assessment of the relationship between the query and the individual document.

All of the models described in this paper are attempts to evaluate the performance of an information retrieval system through one of three approaches. Yu-Ku (16) has stated that the cost of storing data must not be too great and has to be justified by the expected net benefit derived from using such information. Goffman (5) has also stated that the function of any information retrieval system is to provide the user with relevant information on request. Thus it is postulated that the criteria used in any evaluation of any retrieval system should be based upon the user's requirements and the cost benefits of the system.

In the following pages, factors which may be used to evaluate the performance of a retrieval system will be discussed. Most of these factors have been discussed earlier in this paper, but these have been deemed to be highly essential in any evaluation of an information system from the user's viewpoint. All of the factors are not mutually independent, but are interrelated. The prime factor in any retrieval system is the budgetary restrictions and is related to the system's cost benefits. This factor affects the

system's performance and the other factors are dependent upon this one. The criteria of the model which is proposed strongly tilt towards the requirements of the user and have been enumerated by Lancaster (10). Such a model can be strongly defended on the grounds that a user doesn't really care or want to know how the elements or components of the system are interconnected, or what type of indexing was used by the administrator or operator, or the various types of search strategies used. Their one criterion is "how well does the system respond to my question or query." Mathematical formulae are not discussed due to a lack of expertise in this area.

The first factor that will be discussed is precision. Precision refers to those references which are presented to the user and are relevant to his purposes. The system must have the ability to screen all documents in the system and judge which documents are suitable for the user's purpose. King and Bryant (9) call this the system relevance response.

A second factor that enters into this model is that of recall. Recall refers to the total relevant references that are produced by the search.

Both of these factors constitute what may be labeled the one big factor of evaluating the performance of an information retrieval system and this is relevance. The concept of relevance is unavoidable if one evaluates a system. Recall and precision ratios are indicative but taken together give a much better criteria from which to evaluate. The concept of relevance in evaluation is controversial (1), but this is puzzling. If a system provides information or references which are irrelevant to the user, then that system is regarded as worthless by that user, regardless of the superiority of the other elements in that system. An analogy would be that if one's mail is continually filed erroneously, the consumer cannot evaluate the Post Office

as excellent, despite the presence of many personnel, pretty vehicles, etc. Goffman (5), however, warns that the relevance of the document depends on the knowledge known at that time--hence relevance is not absolute. One document's information may affect the relevance of the other documents, i.e. document surrogates are not independent. This is correct, however, it is felt that relevance is the prime factor in the user's evaluation of a system.

A third factor of this model is the form of output. The output must be in a form that the user can decipher. Any output must be useful to the user. Lancaster (10) notes that the form has a definite effect on the precision tolerance of the user. The staff operator or administrator should make every effort to provide the user with output in a form that is readily perusable or easily capable of being utilized by the user. If the system has been judged to provide a high degree of relevant references, then the user will have confidence in the output, regardless of the particular format. A fourth factor very closely related to the form of output in this model is the effort required by the user in order to utilize the system. User control is advocated. The system must be flexible and allow the user to work with the system alone. An intermediary is not favored since complete communication is difficult at best and it is felt that the time spent in query negotiation could be better utilized in teaching the user to operate the system. One such system is that of SMART (11) which is an interactive system and uses iterative search strategies which allow the user to select the search options. PLATO is another such system. Both can utilize a cathode ray tube display and thus the output appears on the screen in English and the user can select and evaluate any references which are retrieved.

A fifth factor is the response time of the system. In any system the delay between the query input and output is critical. If batch posting is utilized, the user can suffer an inordinate delay which may affect the value of the output.

Lancaster (11) also includes coverage as a factor, but this has also been discussed by Hjerppe and Lindqvist (6). Coverage will undoubtedly affect the user's evaluation of the system, but this is primarily in the domain of the staff operator or administrator and is related to the cost benefits of the system.

Cost benefits as defined by Lancaster and Fayen (11) is the relationship between the benefits of the system and the costs of providing it. Frequency of usage in a field can be easily determined by the staff operator or administrator and if the number warrants, information or references can be entered into the system. The factor of coverage then is a decision made by the administrator based on the number and frequency of usage. A system may have few entries in one field where users have demonstrated little usage, and many entries in another field which has a high number of users. As was stated previously, the overriding factor which affects the performance of a system is the budgetary restriction. The administrator, by surveying the frequency and number of users, can wisely decide which fields or areas are to be covered in the system and thus increase the cost benefits of the system and satisfy user requirements.

The user requirement factors and the cost benefit factor can be utilized to satisfactorily evaluate the performance of an information retrieval system. In the end, however, it is the user who must judge the system for it is he who must be satisfied with the system's final product.

BIBLIOGRAPHY

1. Annual Review of Information Science and Technology. (Editor: C. A. Cuador). Vol. 3. Chicago, Ill.: Encyclopedia Britannica, Inc., 1968.

2. Cigler, K. "Simulation Methods Applied In Designing and Information Systems." Information Storage and Retrieval. 6(4). Oct, 1970. pp. 307-312.

3. Cooper, M. D. "A Cost Model For Evaluating Information Retrieval Systems." Americal Society for Information Science Journal. 23(5). pp. 306-312.

4. Cooper, M. D. "A Simulation Model of an Information Retrieval System." Information Storage and Retrieval. 9(1). 1973. pp. 13-32.

5. Goffman, W. "An Indirect Method of Information Retrieval." Information Storage and Retrieval. 4(4). 1968. pp. 361-373.

6. Hjerppe, E. R. and Lindqvist, M. G. "A Model for the Evaluation of Information Retrieval Systems Considering the Decision Situation of the User." Proceedings of the 34th Annual Meeting of the American Society for Information Science. Vol. 8. 1971. pp. 77-81.

7. Keith, N. R., Jr. "A General Evaluation Model for an Information Storage and Retrieval System." American Society for Information Science Journal. 21(4). pp. 237-239.

8. King, D. W. and Bryant, E. C. "A Diagnostic Model for Evaluating Retrospective Search Systems." Information Storage and Retrieval. 6(3). July, 1970. pp. 261-272.

9. King, D. W. and Bryant, E. C. The Evaluation of Information Services and Products. Washington, D. C.: Information Resources Press, 1971. 303pp.

10. Lancaster, F. W. Information Retrieval Systems. New York: John Wiley and Sons, Inc., 1968. 217pp.

11. Lancaster, F. W. and Fayen, E. G. Information Retrieval On-Line. Los Angeles, Cal.: Melville Publishing Co., 1973. 587pp.

12. Papier, L. S. "Criteria and Protocol For Immediate Evaluation of Information Systems." Proceedings of the 34th Annual Meeting of the American Society for Information Science. 1971. Vol. 8. pp. 83-90.

13. Rettenmayer, J. W. "File Ordering and Retrieval Costs." Information Storage and Retrieval. 8(2). 1972. pp. 79-95.

14. Rzasa, P. V. and Baker, N. R. "Measures of Effectiveness for an University Library." American Society for Information Science Journal. 23(4). pp. 248-253.

15. Williams, W. F. Principles of Automated Information Retrieval. Elmhurst, Ill.: The Business Press, 1968. 468pp.

16. Yu-Ku, Li. "Evaluating Information Storage and Retrieval Systems--A Decision Theory Approach." Information Storage and Retrieval. 9(5). 1973. pp. 281-291.